Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:51:27
REFERENCE NO: 253

Author Names & Affiliations

- Shucheng Yu - Associate Professor in Computer Science ,University of Arkansas at Little Rock
- Mariofanna Milanova - Professor in Computer Science, University of Arkansas at Little Rock

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Machine Learning, Computer Vision, Cloud Computing and Big Data, Applied Cryptography, Security and Privacy in Smart Systems, and Wireless Networks

Title of Submission

Cloud Multimedia: Real - Time Text, Face and Gesture Recognition using a Mobile -Cloudlet - Cloud Architecture

Abstract (maximum ~200 words).

The goal of this project is to develop a framework and infrastructure to allow real-time analyze of multimedia data. Currently we are working on the projects related to security and augmented reality. The future surveillance solutions will mix facial recognition with other branches such as voice and gesture recognition, biometrics, sentiment analysis, and social relation graph analytics to better detect security threads. Recently Deep Convolutional Neural Network (CNN), as the cutting edge Machine Learning technology, has gained a lot of interest in the areas of computer vision.
Despite its high expressiveness, CNN generally suffers from intensiveness of computation. A practical deep Neural Network can possibly include a large (sometimes millions) of parameters to train and evaluate. This could be a matter of concern in working with large scale data and/or being in real-time regimes. Another still unsolved question is: How close the convolution nets are to human visual processing? In this study biological motivated deep learning architecture is proposed. New architecture consists of model of Visual Attention and Tensor Decomposition based on the Radix (2x2) Hierarchical Singular Value Decomposition. Proposed tensor decomposition simulates 2x2 receptive fields.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The main objectives of this study are:
1. To develop deep learning convolution network that comprises visual attention model. This will help localizing objects in the scene.
2. To develop deep learning convolution network including the called Radix-($2\times^{2}$) Hierarchical SVD (HSVD). This will improve locality, or

more specifically mimic 2x2 receptive fields.

Face and motion recognition is an important area of computer vision with rich literature and diverse application in neurosciences, security and medical sciences. The future surveillance solutions will mix facial recognition with other branches such as voice and gesture recognition, biometrics, sentiment analysis, and social relation graph analytics to better detect security threads. Face recognition is a complex process. The slight visual details of human faces such as orientation, lightning condition, etc, pose a vast amount of complication to the problem of face recognition. Recent studies show that deep neural network develops representations that challenge to mimic mammalian neocortex [2] or even human brain [3]. Regardless of CNN superior object categorization abilities, ConvNets show rather poor object localization results, with the top-performing model (GoogLeNet) in the ImageNet Large Scale Visual Recognition Challenge 2014 scoring 93.3% on a single object categorization task, yet localizing that object with only 73.6% accuracy [1],[4]. Further research in computer vision and machine learning significantly dependent nor only on speed improvement of the algorithms but also includes more studies matching human perception.

To the best of our knowledge there are few studies incorporating visual attention and ConvNet [5]. Humans are capable to detect quickly area (object) of interest in clutter image. Humans don’t see entire image as a static representation, attention machine incorporate salient features and allows region of interest to come to the forefront as needed. Visual attention is measured with eye gaze position. In [6] multi-scale, sliding window approach is used for object localization and detection in ConvNet.

The proposed architecture integrates the concept of visual attention model in deep learning architecture.

Our attention model is based on cognitive signal processing cycle including three major components: feedback, learning and information preservation. To guarantee information preservation we propose to use layered signal decomposition using unsupervised learning to learn and adjust both local and global filters. The learning process is fast because the feature selection algorithm is based on the statistical properties of the scene. For example, in images the closer pixels are likely correlated .The proposed system extracts contextual information. The new approach is based on the hypothesis that human attention is using consecutive approximations with increasing resolution for selected regions of interest. The system also blends deep learning with reinforcement learning. The decision maker (human or machine) autonomously takes actions to achieve the formulated goals. For example, in images two identical patterns allocated in different locations in the image usually express the same semantic content. The new features at the proposed architecture are: locality, sharing and pooling.

We already proved the efficiency of our visual attention model for shallow (single layer) Neural Network . [7] [8]

For the experiment the designed stimulus sets will be approve by our collaborators “psychophysics” from UAMS. The eye movements will be recorded using our eye tracking system .

Recognizing and simulating human behavior has also many applications in the fields of Biomedical information (BI) and the Health Sciences. In 2004 a joint Bioinformatics Program was created between UALR and UAMS. Two years ago, in 2015 a new department Biomedical Information (BI) Department was created at UAMS. Our team began collaborative work with BI analyzing EEG and fMRI images. We rent from UAMS workstation with GPU and installed DIGITS NVIDIA. Using NVIDA digits we obtained the 98% and 99% classification accuracy .The time needed to get the results was a second’s .The preliminary results were presented in IEEE conferences[15], [16].

For Objective2: There are evidences that in ConvNet the deeper one goes, the better the performance will be . The Shallow network (NN with only one hidden layer) corresponds to CANDECOMP/PARAFAC decomposition (CP rank-1) decomposition, but deep network corresponds to Hierarchical Tucker decomposition.

To the best of our knowledge all tensors represented by Hierarchical Tucker (HT) decomposition cannot be efficiently realized by the classic CP (rank-1) decomposition. The problem of depth efficiency is discussed by [9]. Cohen and Shashua in [10] proposed convolutional arithmetic circuits to resolve depth efficiency problem in popular convolutional rectifier networks . They implemented 1x1 operators factorize tensors in SimNets .

The basic methods for tensor decomposition [11] are higher-order extensions of the matrix SVD: the CANDECOMP/PARAFAC (CP) which decomposes the tensor as a sum of rank-one tensors, and the Tucker decomposition, which is a higher-order form of the Principal Component Analysis (PCA).

We proposed new Tensor Decomposition based on the Radix (2x2) Hierarchical Singular Value Decomposition called Radix-($2\times^2$) Hierarchical SVD (HSVD)[12], [13], which to replace the famous Multilinear Singular Value Decomposition (MSVD) [14].

The proposed study will help answering standing question in the field: How to improve speed and accuracy for object, pattern categorization.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

It would be good to know what the data movement scenarios are. For example, picture is taken with an iPad and it is send to be identified, etc. If we take a surveillance camera, 960 hours of surveillance would require about 500GB. That's about 40 days processing time. If there were 100 cameras it would be 5 TB required for 40 days. Then you need space for extracting frames, classifiers, for recognition, etc.
On the other hand, modern distributed processing technologies, such as NoSQL solutions, offers practical possibilities to deal with the complexity of massive scale data processing pipelines. The capability of horizontal scalability, which these technologies offer, results in direct relationship of computational power to the available cluster resources. In the specific case of a Deep Learning, a more powerful cluster can lead to a larger Neural Network, hence a better accuracy of prediction, lower training time and a more near to real-time prediction process.
Due to the current volume, we are unable to store all the available data in our cluster. Besides the data is spread out throughout local departmental data centers and processing it takes a massive inter data center network communications delays. Being able to digest more data and having a more powerful network infrastructure in conjunction with exploiting NoSQL technologies would enable us to have a stronger computational platform in order to design better algorithmic solutions to the problems that we are working on. A direct outcome of these would be more accurate statistical and machine learning modeling.

• Existing cyberinfrastructure:
The Computational Research Center (CRC) at the University of Arkansas at Little Rock hosts the most powerful computing instruments on campus. It serves as the hub for various research projects at UA Little Rock spanning areas such as bioinformatics, social networks, computational chemistry and physics. The CRC instruments are featured with massive distributed computing resources including general-purpose CPUs, RAM and secondary storage. For example, its Rocks 5.4 cluster consists of 64 Dell PowerEdge machines (each with 8 Xeon processors and 16GB Ram, for a total of 512 cores), 4TB storage, Gigabit Ethernet, and Infiniband interconnection among computing nodes for minimal IPC latency as well as between computing nodes and a 40TB Lustre parallel file system for fast file access. Erbium, our big memory machine by HP, has 80 processors (160 hyperthreaded) and 4TB memory in a single node.

• Limitations:
The CRC instruments can perfectly fulfill the requirements of most scientific computing tasks which demand large-scale parallel processing over general-purpose CPUs. However, they are extremely limited in processing data-intensive tasks such as image processing in AR/VR and machine learning algorithms. The single-node machine Erbium is capable of handling these types of tasks but this instrument is usually overwhelmed while serving ALL the research projects on campus. While GPUs or GPU clusters have been pervasively adopted for aforementioned applications, our CRC instruments have NOT yet equipped with any GPU. This unbalanced configuration significant constrains the full utilization of this highly distributed computing infrastructure.

• Needs:
To compensate the limitations of the existing CRC instruments, at least 8 computing servers with the following configuration are necessary:
• CPU – 4 Xeon processors
• GPUs - 4 NVIDIA TESLA K80
• System Memory - 64 GB DDR System Memory
• Storage - 1 TB SATA SSD + 3 TB 7200 rpm HDD for Long-term Data Storage

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

• New infrastructure will help increasing collaborative research between UALR and industry.
• New infrastructure will help students to create new project in the following classes : Machine Learning , Data Mining , Artificial Intelligence, Image Processing, Computer Graphics , Sensor Networks, Telecom/ Networking
• Students will learn new concepts and become ready for new challenges in the future .

References:
1. Google apologizes for Photos app’s racist blunder [Internet]. BBC News. [cited 2015 Nov 28]. Available from: http://www.bbc.com/news/technology-33347866
2. Kubilius J, Bracci S, Op de Beeck HP, Deep Neural Networks as a Computational Model for Human Shape Sensitivity , Journal of Vision, September 2016.
3. Khaligh-Razavi S-M, Kriegeskorte N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Comput Biol. 2014 Nov 6;10(11).
4. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. ArXiv14090575 Cs [Internet]. 2014 Sep 1 Available from: http://arxiv.org/abs/1409.0575
5. Xu K and ets , Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, csLG , Apr 2016
6. P Sermanet, D Eigen, X Zhang, M Mathieu, R Fergus, Y LeCun, OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks, International Conference on Learning Representations (ICLR 2014), 16
7. Milanova M. , Mendi, E, Attention in Image Sequences: Biology, Computational Models, and Applications, Chapter 6, Advances in Reasoning-Based Image Processing Intelligent Systems, 2012, pp147-170.
8. Milanova M., Rubim S., Kountchev R., Kountcheva R., Combined Visaul Attention Model for Video Sequences , ICPR 2008
9. Cohen N., Sharir, Shashua A., On the Expressive Power of Deep Learning: A Tensor Analysis, csNE, 27 May , 2016 .
10. Cohen N., Shashua A , Convolutional rectifier networks as generalized tensor decomposition, ICML 2016.
11. A. Cichocki, D. Mandic, A. Phan, C. Caiafa, G. Zhou, Q. Zhao, L. De Lathauwer, Tensor Decompositions for Signal Processing Applications, IEEE Signal Processing Magazine, 32 (2), 2015, pp. 145-163.
12. R. Kountchev, R. Kountcheva, Radix-(2×^{2}) Hierarchical SVD for multi-dimensional images, Proc. of the IEEE Intern. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS’15), Nis, Serbia, Oct. 14-17, 2015, pp. 45-55.
13. R. Kountchev, R. Kountcheva, New approaches for hierarchical image decomposition, based on IDP, SVD, PCA and KPCA, Chapter 1 in book “New Approaches in Intelligent Image Analysis: Techniques, Methodologies and Applications”, May 2016, pp. 1-58.
14. H. Lu, K. Plataniotis, A. Venetsanopoulos, MPCA: Multilinear Principal Component Analysis of Tensor Objects, IEEE Transactions on Neural Networks, Vol. 19, No 1, January 2008, pp. 18-39.
15. A. Taqi, F. Al_Azzo, M. Milanova, Classification and Discrimination of Focal and Non-focal EEG signals Based on Deep Neural Network, IEEE International Conference on Current Research in Computer Science and Information Technology (ICCIT-2017) (accepted)
16. Human Actions Recognition Based on 3D Deep Neural Network IEEE Conference on Mew Trends in Information and Communication Technology Applications, 2017 , (accepted).

Consent Statement